TRial of Aggregate Data Exchange for Maintenance of certification and Raising Quality (TRADEMaRQ)

Project Period: September 30, 2014 – December 31, 2017
This project was supported by grant number R21HS022583 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

AHRQ Project Officer: Bryan B. Kim, PhD

Robert L. Phillips, Jr. MD MSPH
American Board of Family Medicine
1648 McGrathiana Pkwy, Suite 550
Lexington, KY 40511
859-269-5626 x 1253

Lars Peterson, MD PhD
American Board of Family Medicine
1648 McGrathiana Pkwy, Suite 550
Lexington, KY 40511

Ted E. Palen, PhD, MD
Colorado Permanente Medical Group
Institute for Health Research
2550 S Parker Rd, Suite 200 Aurora, CO 80207

Scott A. Fields, MD, MHA
OCHIN
1881 SW Naito Parkway Portland, OR 97201

Michael L. Parchman, MD, MPH
Kaiser Permanente of Washington Health Research Institute
1730 Minor Ave, Ste 1600 Seattle, WA 98101

John Johannides, PMP
American Board of Family Medicine
1648 McGrathiana Pkwy, Suite 550
Lexington, KY 40511

Abstract:

Purpose: TRADEMaRQ tested whether quality reporting could be a byproduct of clinical care, and whether reducing reporting burden for Maintenance of Certification could influence more meaningful improvement efforts. The study had three aims: 1) To test the capacity for exchange of whole-panel, standardized quality measures from clinical networks; 2) To study whether viewing quality measures and peer comparisons would affect quality improvement efforts selected; and, 3) To study whether viewing quality measures and peer comparison would influence change in outcomes after quality improvement efforts.

Methods: We conducted a cluster randomized controlled trial with 4 partner organizations to present quality data through each physician's online certification portfolio prior to certification activity choice. We assessed measure and transfer errors, choices of quality improvement relative to quality gaps, and change in quality.

Results: Health systems varied considerably by patient demographics and payer mix. Of 2,570 eligible physicians, 258 participated. Of 19 measures negotiated for use, five were used by all systems. There were more than 15 identified errors including data delivery, measure modifications, and nonsensical measure results. Despite lengthening the duration of the trial, nearly 20% of the sample did not complete a certification activity and nearly 25% of physicians in the intervention arm never reviewed their quality dashboard. Only 27% of improvement activities were completed in an area where the physician's quality was below average, and there was no difference between intervention and control groups. There was likewise no significant difference in quality improvement between the two arms.

Key Words: Quality Improvement, Certification, Primary Care, Quality Measurement

Purpose:

The objective of the TRADEMARQ study was to make quality reporting a byproduct of ambulatory care and ongoing quality improvement. It had three aims: 1) To test the capacity for exchange of whole-panel, family physician quality measures from clinical networks; 2) To study whether viewing quality measures and comparison to peers would affect the types of self-assessment modules and quality improvement efforts that were chosen by family physicians (randomized); 3) To study whether viewing quality measures and comparison to peers would influence the degree of change in outcomes after quality improvement efforts.

We have previously published research demonstrating relationships between participation in continuous certification, particularly quality improvement activities, with improved quality, particularly if combined with federal quality reporting (Physician Quality Reporting Service). At a fundamental level, we sought to reduce redundancy and burden of measure reporting by creating an automated path for their quality measures to flow into the certification process from their home clinic networks—something that NCQA is also now piloting for PCMH certification.

Scope:

Most Family Physicians are now participating in Maintenance of Certification (MOC), measuring and reporting quality measures from limited patient samples before and after a quality improvement effort (Part IV requirement). As ABFM sought to take next steps towards broader and more regular assessment of quality, and to simultaneously reduce reporting burden, we wanted to test whether such efforts would improve outcomes. This intention was consistent with AHRQ's aims for HIT and Quality and with those of the Office of the National Coordinator for Health Information Technology (Direct Project and Nationwide Health Information Network Exchange).

We developed partnerships with four clinical systems for this study: Kaiser Permanente Colorado, OCHIN, South East Texas Medical Associates (SETMA), and Kaiser Permanente Washington (formerly Group Health Cooperative). All four use standardized, physician-level quality measures and were willing to test ways to securely share these. We developed project leadership and secure data exchange pathway to all four entities to iteratively test means of automating direct transmission of Family Physician measures, exploring both technical and legal solutions (Aim1). For Aim 2, physicians were randomized within their system for exposure to their measures and comparisons to peers before or after they choose their self-assessment and quality improvement projects. Randomization was clustered by clinic to avoid risk of contamination. Within each randomization arm, we conducted formal correlation analysis between Self-Assessment Modules (SAM) and Part IV selection and quality measures using hierarchical logistic regression. For Aim 3, we assessed changes in crude rates of quality measures from baseline to study completion using a difference in difference analysis which controls for physician characteristics and a longitudinal hierarchical logistic regression model fit to account for dependence of observations.

Methods:

Study design

TradeMarq was a practice level cluster randomized trial.

Data sources/collection

A secure file transfer protocol (SFTP) server was developed to support measure data transfer from each entity to the ABFM on a biweekly basis. The eligible e-Certified Quality Measures (eCQMS) were negotiated with the goal of having standardized, common data across systems to avoid risk of differences in calculation. We also collected routine data from the ABFM certification portfolios of enrolled physicians, including views of their portfolios, views of measure dashboards, time spent reviewing measures, selection of SAMs and IAs, and changes in quality measures over time.

Interventions

The intervention was delivered via the physician's ABFM portfolio. The ABFM created a dashboard of quality data reported by the partners for the physician, with a comparison to the average quality for the measure of all physicians in the study. Physicians in the intervention arm were presented with their dashboard prior to being able to select an ABFM certification activity. Physicians in the control arm were shown their quality dashboard after selecting a certification activity. Both groups of physicians were able to use the data from the quality

dashboard as their quality measures for the quality improvement activity, if done through the TRADEMaRQ module which was automated for enrolled physicians and presented to them as they logged into their portfolio.

Measures

Overlap in Quality Measures Used in TRADEMaRQ by Partner

TRADEMaRQ Measures NQF # (description)	KPWA (Group Health)	KPCO	OCHIN	SETMA
0018 (high blood pressure)	X	X	X	X
0028 (smoking cessation)		X	X	
0031 (breast cancer screening)	X	X	X	X
0032 (cervical cancer screening)	X	X	X	X
0034 (colorectal cancer screening)	X	X	X	X
0041 (influenza immunization)			X	X
0043 (pneumonia vaccination – adults)			X	X
0055 (diabetes – eye exam)	X	X	X	X
0056 (diabetes – foot exam)		X	X	X
0059 (diabetes – hemoglobin A1c)		X	X	X
0062 (diabetes – urine protein screening)	X		X	X
0064 (diabetes – LDL)			X	X
0068 (IVD: aspirin or another antithrombotic use)		X	X	
0070 (CAD: beta-blocker w/ prior MI or LVSD)	X		X	
0075 (IVD: lipid panel and LDL control)		X	X	
0081 (HF: ACE or ARB)	X	X	X	
0083 (HF: beta-blocker for LVSD)			X	
0418 (depression screening)			X	
0419 (current medication documentation)			X	X

Limitations

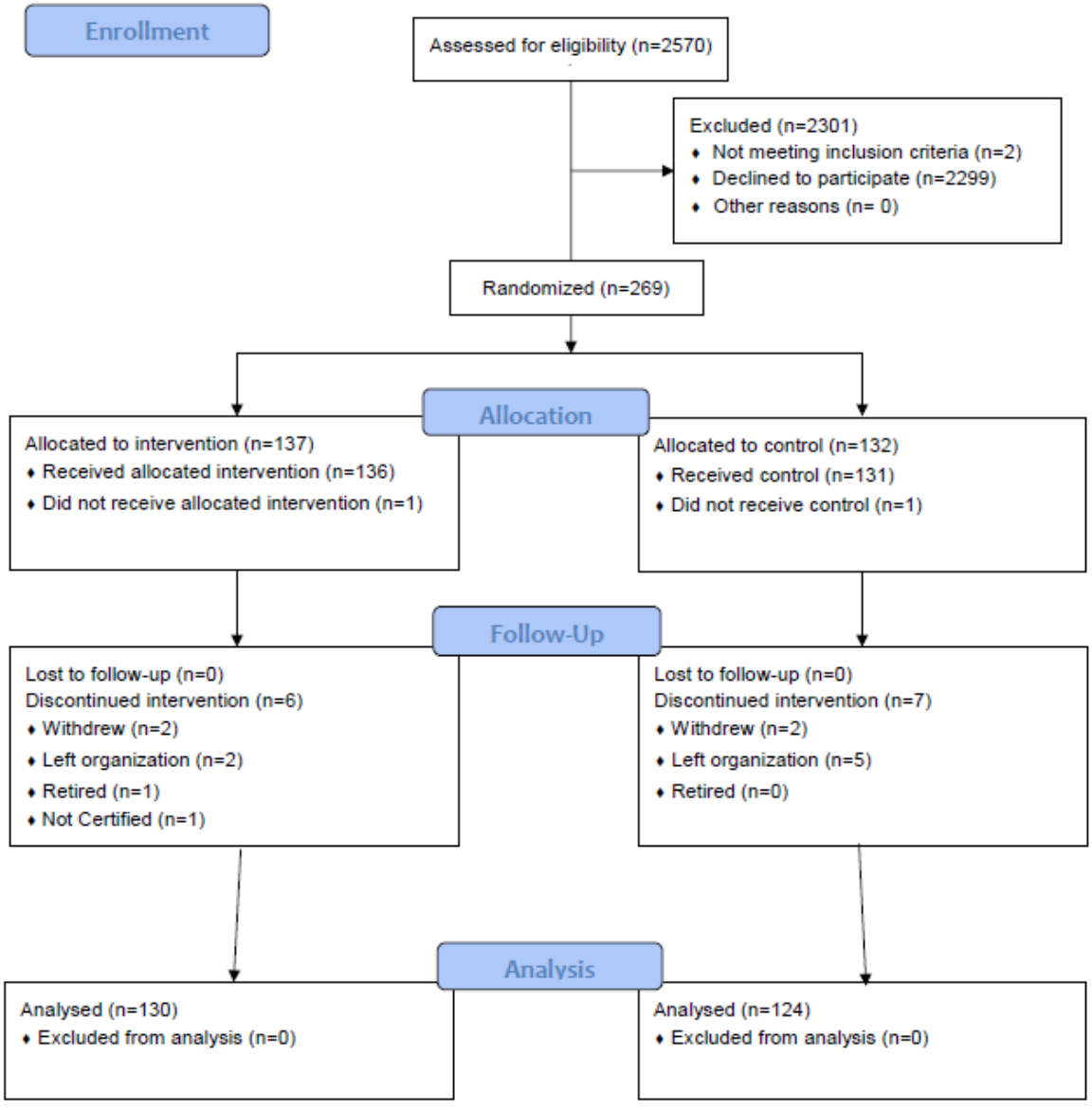
Prevention of contamination in SETMA was potentially compromised by the fact that this system reports clinician-level measures on their website and use them routinely for internal assessment, coaching, and improvement. This concern was substantiated by the fact that all of the SETMA clinicians were more likely to choose improvement activities associated with their measured care gaps and were more likely to improve on those measures over time.

TRADEMaRQ did not achieve the enrollment needed to achieve pre-assessed power potentially producing a Type 2 error.

Results

We identified 2,570 possible participants at the 4 partner organizations and invited them to participate. Two of these did not meet inclusion criteria and 2,299 declined to participate. We randomized 269 physicians in 6 waves of enrollment at the practice level to the intervention or control arm (Figure 1). All participants received the intervention. Thirteen participants left the study with a majority of these leaving the partner organization (n=7). Only 4 participants withdrew from the study. Our final analytic sample was 254 physicians; 130 in the intervention arm and 124 in the control arm.

Figure 1: Flow of Participants through Study



Principal findings

Our final sample had a mean age of approximately 50 years, was slightly more than half female, were almost all allopathically trained (MDs) and US medical graduates. There were no differences in demographic characteristics or medical knowledge as measured by their most recent ABFM Family Medicine Certification Examination score (table 1). There were at least 20 discovered errors in data calculation or transmission errors (table 2).

Table 1. Demographics of TradeMarq Participants in Final Analytic Sample

	Overall (n=254)	Test Group (n=130)	Control Group (n=124)	p-value
Physician Characteristics				
Age in Years, Tertile				0.31
<43.49 years (N=84)	39 (2.9)	39.3 (2.7)	38.8 (3.1)	
43.49-55.0 (N=87)	48.3 (3.3)	48.2 (3.3)	48.5 (3.3)	
>55.0 (N=83)	62.4 (4.3)	63 (4.5)	61.8 (4)	
Female Gender	143 (56.3%)	74 (56.9%)	69 (55.7%)	0.94
MD degree (vs. DO degree)	236 (92.9%)	124 (95.4%)	112 (90.3%)	0.12
International Medical Graduate	19 (7.5%)	7 (5.4%)	12 (9.7%)	0.20
Most Recent ABFM Certification Examination Score, Tertile				0.76
Score <=530 (N=94)	470.4 (44.8)	468.8 (44.8)	472.3 (45.3)	
Score 530-620 (N=82)	583.3 (24.2)	586.1 (25.6)	580.9 (22.9)	
Score >620 (N=78)	686.3 (48.8)	688.1 (49)	684.2 (49.1)	
Organization N (%)				0.096
KPWA	90 (35.4%)	54 (41.5%)	36 (29%)	
KPCO	64 (25.2%)	26 (20%)	38 (30.7%)	
OCHIN	94 (37%)	46 (35.4%)	48 (38.7%)	
SETMA	6 (2.4%)	4 (3.1%)	2 (1.6%)	

Table 2: Quality Measure Data Errors

Error types	Examples	Fix
Measure miscalculation	• Incorrect numerator or denominator • Incorrect data period (measurement period required 12 months, but 11 months used) • Incorrect denominator inclusion criteria used for greater than 1 year • Numerator > denominator error • Patient panel (erroneously) reduced to zero • Physician moved clinics and changed panels so that measures could not be reconciled; removed from the study • Significant change in scores for 5 measures • Third party vendor managing measures was only producing measures for system monthly and was incorrectly compressing for biweekly reporting so study measures/panel size did not jibe with system measures/panel size (physicians complained)	Corrected data sent and uploaded manually after manual removal of incorrect data Revised measure calculations Error caught internally and repaired; delayed transmission Fixed reporting period compression error
Data Delivery error	• Delay in data delivery • Blank file sent • Incorrect NQF number attached to file	Corrected, resent, manual data replacement
Non-enrolled physician data sent	• Physician data sent before they were enrolled/randomized; • Ineligible physician data sent	Physician enrolled and randomized or excluded
Data reporting interrupted	• Physician data reported for one period but not another; • Internal system change caused a measure to not get reported • Source databased moved and transmission credentials not configured • Critical subsystem source failure, 6-week delay	Updated files sent and manually uploaded
Host receiving server not running	System update interruption	Server brought back on-line
Third party errors	• Two years into study, learned that a third-party company was doing measure management and transmitting incorrectly • Third-party processes caused several month delays in file transmission around turn of calendar year	Worked directly with vendor to correct calculation or transmission errors Files caught up once data sent by third party

Table 3: Number of Certification Modules Completed by Study Participants

	Overall (n=254)	Test Group (n=130)	Control Group (n=124)	p-value
No module completed	50 (19.7%)	32 (24.6%)	18 (14.5%)	0.043
At least one module completed	204 (80.3%)	98 (75.4%)	106 (85.5%)	
Median (IQR) number of modules completed for those completing at least one module	4 (3-6)	4 (2-6)	5 (3-6)	0.0023

Table 4. Proportion of physicians in the Test Group who Interacted with the Quality Dashboard and the mean time spent reviewing the Dashboard in the 2 weeks period before starting a Certification Activity (n=130)

	Overall (n=130)	SETMA (n=4)	KPCO (n=26)	KPWA (n=54)	OCHIN (n=48)
Never looked at Quality Dashboard	24	1	9	9	5
Looked at Quality Dashboard 2 weeks prior to Certification Activity	59	0	9	23	27
Median seconds (IQR) reviewing dashboard 2 weeks prior to Activity	82 (52-155)	0 (0-0)	75 (43-179)	86 (55-131)	84.5 (53-156)

Table 5. Proportion of Physicians Choosing to Work on Certification Activities with Below Average Quality Measures by Organization and Intervention Arm (Good Choice)

	Certification Activity Reflects Low Quality (N=186)	Test Group	Control Group	P-value
KPWA	13.6 (28.3)	17.7 (34.1)	6.4 (10.2)	0.86
KPCO	22.0 (19.5)	25.2 (23.5)	20.1 (16.7)	0.59
OCHIN	38.4 (29.4)	40.6 (28.9)	36.6 (30.0)	0.52
SETMA*	57.9 (7.1)	50 (.)	61.8 (2.6)	.
Total	27.2 (28.5)	28.8 (31.1)	25.7 (25.9)	0.94

* There was only 1 test participant from SETMA so there is no variation around the proportion and no statistical test.

Table 6. Average Quality Performance for First Measurement and Last Measurement Point by Study Arm

TRADEMaRQ Measures NQF # (description)	Test Group		P-value	Control Group	
	First time Point	Last Time Point		First time Point	Last Time Point
0018 (high blood pressure)	0.68(0.19)	0.72(0.26)		0.74(0.12)	0.74(0.19)
0028 (smoking cessation)	0.24(0.08)	0.19(0.07)		0.25(0.08)	0.21(0.08)
0031 (breast cancer screening)	0.82(0.09)	0.73(0.14)		0.82(0.08)	0.71(0.16)
0032 (cervical cancer screening)	0.70(0.16)	0.73(0.21)		0.69(0.13)	0.73(0.19)
0034 (colorectal cancer screening)	0.66(0.20)	0.66(0.23)		0.65(0.20)	0.67(0.23)
0041 (influenza immunization)	0.61(0.13)	0.36(0.13)		0.63(0.11)	0.30(0.29)
0043 (pneumonia vaccination – adults)	0.77(0.12)	0.78(0.11)		0.76(0.13)	0.89(0.18)
0055 (diabetes – eye exam)	0.73(0.20)	0.78(0.18)		0.77(0.21)	0.83(0.19)
0056 (diabetes – foot exam)	0.94(0.07)	0.94(0.06)		0.95(0.03)	0.93(0.06)
0059 (diabetes – hemoglobin A1c)	0.51(0.30)	0.58(0.32)		0.56(0.3)	0.57(0.31)
0062 (diabetes – urine protein screening)	0.92(0.10)	0.95(0.06)		0.94(0.05)	0.94(0.16)
0064 (diabetes – LDL)	0.91(0.03)	0.94(0.02)		0.91(0.06)	0.92(0.07)
0068 (IVD: aspirin or another antithrombotic use)	0.83(0.18)	0.85(0.21)		0.78(0.24)	0.89(0.17)
0070 (CAD: beta-blocker w/ prior MI or LVSD)	0.87(0.28)	0.89(0.27)		0.97(0.17)	0.99(0.04)
0075 (IVD: lipid panel and LDL control)	0.83(0.06)	0.86(0.05)		0.82(0.07)	0.86(0.05)
0081 (HF: ACE or ARB)	0.74(0.25)	0.73(0.25)		0.82(0.19)	0.78(0.14)
0083 (HF: beta-blocker for LVSD)					
0418 (depression screening)	0.61(0.22)	0.55(0.28)		0.49(0.26)	0.58(0.25)
0419 (current medication documentation)	0.86(0.12)	0.95(0.03)		0.99(0.00)	0.66(0.57)

Outcomes

TRADEMaRQ experienced low uptake by family physicians in three of the four health systems despite employing several strategies to help them understand the potential value for reducing MOC complexities, and to reduce fear of identification or other potential, perceived risk. Per Aim 1, we experienced breakdowns in effective quality measure data calculation, transmission, or both—except for one health system out of the four. For Aim 2, Nearly 1/5th of the sample did not complete a certification activity despite this being a requirement for MOC, those who did

conducted more than they would typically be required to (table 3), and of those in the test group who did, 2/3rds did not look at their measure dashboard populated with their own measure data (table 4). Also related to Aim 2, most physicians in only one of the four participating organizations chose improvement activities related to their quality gaps (table 5). Per Aim 3, there were no significant differences in quality change between the intervention and control arms (table 6).

Discussion

The low participation and threat to power for assessing meaningful differences between the RCT arms of the trial is a disappointment. Enrollment was voluntary and anonymous to the study PI and administrators to avoid any sense of coercion; however, local collaborators were provided with ample recruiting materials and eligible family physicians had several exposures to email and face-to-face invitations and informed consent. Despite the option to reduce burden for reporting quality measures, to use whole-panel data rather than hand-entered samples, and the ability to combine their actual quality improvement efforts with MOC, we were unable to recruit the sample we anticipated. Regarding Aim 1, the lesson we take away from this is that even mature health systems with long-standing quality assessment functions are prone to breaking down unless quality is routinely used to guide improvement. It may be that if quality measures aren't routinely reviewed or used, errors aren't noticed. It may be that they aren't believed to begin with, errors aren't found or fixed. The clinical system that experienced no errors in quality measurement or problems with transmission, SETMA, is the same one that routinely published clinicians' measures on their website. Measure transparency to that degree may push regular review and intention to make sure they are correct. This may be one of the most important hypotheses to result from this study. Clearly many participants did look at their quality measures in their MOC portfolio, and many chose to work on their weaknesses, but it is hard to understand why some people who, despite the burden reduction of having their whole-panel measures available to them, did not look at their measures before choosing quality improvement activities. The third aim was dependent on the first two, so it is not surprising that there was a lack of difference in measure improvement.

Conclusions

While quality measurement has long been a function in many mature health systems and the Quality Payment Program is likely to increase measure production and reporting, we found little to support our a priori belief that these measures are regularly tracked, fed-back, or reflected on by systems or clinicians. That we discovered multiple errors before they were discovered by the participating organizations is telling. Only SETMA had no discovered errors, and they routinely post measures for all clinicians on their website in full transparency. This level of scrutiny may be required for maintaining measure accuracy. Likewise, we found little in the way of quality review (MOC dashboard review) or targeted improvement (Practice Improvement Activity selection), except for SETMA for which clinicians in both arms were unlikely to view their dashboards but were more likely to work towards improving their gaps (presumably viewing their measures in their own process made looking at their MOC dashboard irrelevant). These findings, while limited to four mature clinical organizations, suggest that the Quality Payment Program is unlikely to meet expectations unless quality measurement becomes a routine part of care and practice improvement. Annual reporting is unlikely to drive routine, regular review and improvement.

Significance

Given the purposeful shift to value-based payments by many payers, the findings from this study should reduce expectations that value or quality measurement will affect clinician awareness or care-related behaviors. It is true that value-based payment raises the stakes for regular quality review and targeted improvement activities, but the processes, culture, and assistance needed are not in place for most systems. Most practices or health systems may need more support in understanding how to integrate routine quality review and improvement.

Implications

Value-based payments are likely to see slow progress in quality and value improvement unless practices and health systems make or find the resources to help with related culture and process change.

List of publications and projects

DeMaio S. Simplifying MOC for Famly Physicians: The TRADEMaRQ Study. *NEJM Knowledge+.* 2015; https://knowledgeplus.nejm.org/blog/simplifying-moc-for-family-physicians-the-trademarq-study/.

Tachibana, C. How can we streamline board certification for family physicians? Group Health Research Institute News and Events. February 5, 2105.

Phillips, R. ABFM to Simplify MOC for Family Physicians and Make It More Meaningful: A Family Medicine Registry. JABFM. 2015; 28(3): 431-433.

Under review:
Phillips, RL, et al. Reliable Measure Exchange Is Not Easy: The Trial of Aggregate Data Exchange for Maintenance of certification and Raising Quality (TRADEMaRQ) Study. JAMIA

Peterson, LE, et al. Physicians Choice of Board Certification Activity is Unaffected by Quality of Care: The Trial of Aggregate Data Exchange for Maintenance of certification and Raising Quality (TRADEMaRQ) Study.